Integrated Water Quality and Aquatic Communities Protocol – Wadeable Streams

Standard Operating Procedure (SOP) #19: Quality Assurance Project Plan

Draft Version 1.0

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Version

Consider the following scenario: A driver is pulled over by an officer of the law for speeding. This officer used a radar gun to determine that the driver (maybe you, the reader) was traveling 12 miles per hour over the speed limit. In your defense, you believe that you were not traveling that fast, and that maybe his or her radar gun was not functioning properly. Maybe the radar gun was reading wrong and that the margin of error on the radar was plus or minus 12 miles per hour? All valid concerns, except that the officer can confidently tell you that the radar gun was calibrated the previous day against a known speed, and that the margin of error is documented at plus or minus 2 miles per hour. Not only that, but they have written documentable proof of it, and in sum, you're *busted*. How was the officer able to know all of this? Because of a Quality Assurance Project Plan!

Just as the officer was confident of the radar gun reading, we have to be confident of our own stream measurements. Our path to confidence is also in a Quality Assurance Project Plan.

This SOP details the Quality Assurance Project Plan (QAPP) for this protocol.

The purpose of the QAPP is to ensure that data produced through this protocol is of a known, documentable, and defensible quality. A clearer way of defining the QAPP purpose is to pose example questions that it is supposed to answer:

- If we measure calcium to be 0.8 mg/L (at noon, in July, from a riffle), how do we know that it is actually 0.8 mg/L, and not 0.9 or 0.7 mg/L?
- If we measure calcium at 0.8 mg/L and in the future it is measured at 1.2 mg/L, how do we know that there has been an actual increase in the calcium and not just measurement error?
- If our multiprobe breaks and is replaced, how do we know that probe B (new) is giving comparable results to probe A (old)?
- If we found insect species A in year 1, but not in year 10, is this because the species was locally extirpated, or because of a taxonomic error, or because of change in taxonomy?

- If we are making a probabilistic survey of the Network wadeable streams, how do we know our sample *really* represents the wadeable streams of the Network? If it represents it now, how do we know that it will represent the streams 30 years from now?
- If our data are compared or shared with other data sources, how do we know that our numbers are representative of other programs?
- If we produce data that do not meet certain standards, how are they handled?
- If we are not able to sample all of the sites or all parameters, do we still make inferences to the park units?
- If a field in the database is left blank, how do we know if that data was not collected, did not exist, or was accidentally not entered?
- If measurements made by the Network are used in a court of law, how do we prove that we followed the SOPs to the letter and that the data represents the best measure of "truth" possible?

The answers to these questions are dealt with in two primary methods: documentation and methodology. The documentation includes the most basic form: the current protocol, from the narrative to the last appendix, and of course, the strict adherence to the methods described in the SOPs. The delivery of data must also include documentation. In this form, much of the documentation is metadata provided with the data. It is also in the strict documentation of all events related to this protocol, even though not directly related to the data about wadeable streams (e.g., the documentation of field crew training and calibration events).

To help ensure quality data, methods are implemented to deal with the inevitable change (e.g., new instrumentation, changing analytical chemistry laboratories). Other methods affecting quality data are data validation and verification steps. In addition, the development of a data collection system that incorporates domain values, pick list, and logic checks is important.

Some aspects of this QAPP will come in the form of guidance. When possible, the following steps and methodology should be carried out. **It is through the documentation of errors, variance, etc. (whatever the source) that the quality of the data is known.** Where possible, this should be included with the metadata. For example, with the cumulative bias procedures (described below), it is recommended that a shift in personnel be accompanied by seven overlapping measurements. For a field crew change, with 2 years in between sampling periods, this is untenable. For the Project Lead, who is responsible for training the field crews, it would be wise for the outgoing Project Lead to sample with the incoming Project Lead. However, personal situations can easily prevent this from happening and the cost of processing seven extra macroinvertebrate samples can easily be out of the budget.

Sampling Frame, Sample Size, and Minimum Detectable Differences

Sampling Frame

The sampling frame for this protocol is all wadeable streams in Lassen Volcanic National Park, Crater Lake National Park, Whiskeytown National Recreation Area, and Redwood National and State Parks (the only component of our design with a probabilistic sampling design), with the following criteria:

- Perennial – This selection criterion is applied to remove habitats that are influenced by seasonal desiccation which could mask other stressors of interest. It also ensures that data collection can always occur at the sites, assisting in data completion goals.
- Less than 1000 m from a travelable road or trail – This selection criterion reduces logistical constraints to field crews, such as travel time, to ensure that each site can be sampled in the allotted time frame for achieving sampling objectives.
- In mild topographies with stream slopes <15% – This selection criterion ensures crew safety and that access to streams is doable.

As Irwin (2008) points out, long-term monitoring plans must deal with the possibility of the population of interest "drifting" in and out of the sampling frame. An example of a realistic event causing this is the park making new roads or trails (or decommissioning roads and trails). This would change the list of habitats within our sampling frame.

As part of this protocol, the adequacy of the sampling frame will be revisited on a regular interval of 15 years (every four sampling periods). If the sampling frame or population of interest is found to have changed within this period, corrective action, as necessary to accomplish the goals of this protocol, may be undertaken.

Sample Size and Minimum Detectable Differences

This protocol focuses on park-level inferences, using measures of central tendency to track changes in status and trends. To achieve this, with ample statistical power in select parameters (see narrative), a minimum sample of 25 streams is required. Based on data completeness goals (see below), we have increased the sample size to 30.

Minimum detectable differences are a question of power and are detailed in the power analysis, covered in the protocol narrative.

Data Comparability

Comparability is the measure of confidence that one dataset, element, or method can be considered as similar or identical to another (SWAMP 2008). The goal of this document is to ensure that if the programmatic requirements are fulfilled, then the data collected by this protocol is considered to be a similar quality level.

It is outside the scope of this protocol and Network budget to do comparability studies to other methodologies, be it Forest Service, academic, or state sponsored monitoring. As research or external funding is available to address these comparability studies (and as need dictates), Network staff may work towards these goals. During the development of this protocol, methods and parameters being implemented by other agencies were considered and in some cases applied to this protocol However, the fundamental goal of this QAPP and overall protocol is to ensure comparability *within* our program. The major source of potential variation that could affect data comparability is protocol or equipment changes.

Data Completeness

The Klamath Inventory and Monitoring Network recognizes that a certain percentage of samples of all types (invertebrates, water chemistry, etc.) will fail. Reasons for failing may include lost samples, dehydrated samples, samples that exceed holding times, or sample contamination.

Data completeness goals are based on a multi-step process detailing the minimum sample size needed to make statistical inferences about the population of interest (Irwin 2006). Once this minimal sampling size is calculated (see Sampling Frame and Sample Size and Minimum Detectable Differences, above), the number of samples that will fail is estimated and the sample size is increased by the same percentage. This procedure is complicated for multi-parameter protocols, such as the one here. Water chemistry samples (with multiple types), invertebrate samples, water probes, and GPS files may all fall short of precision targets, but at different failure rates. However, since we are making inferences to a population of streams, the ultimate sample size for this protocol is *the number of sites*.

During the initial scoping period, it was determined that a minimum of 25 samples are needed to characterize the population of streams with desired levels of precision (A. Merton, statistician, personal communication). Following the guidance of Irwin (2008), we increased our sample size to 30 streams reaches to accommodate unforeseen problems.

Cumulative Bias

The term bias has many definitions, even within the realm of statistics. Here, bias is taken to be a systematic error in measurements. Over the length of a monitoring program, bias may cumulate from many sources: collector bias, instrument bias, protocol bias, etc. With the obvious expectation that personnel and gear will change and protocols will be revised over the years (SOP# 24: Revising the Protocol), QAPP methodology for dealing with change and minimizing and documenting the cumulative bias are laid out below. Following these procedures will allow for Data Comparability.

Any change in the following categories will be documented in the metadata produced during the project. Just as field crew personnel changes from year to year are documented, the different laboratories (although changing labs should be a severe option) will also be documented.

Change in Personnel

When possible, it is recommended that personnel changes be accompanied by an overlap of seven measurements. For the streams protocol, this extrapolates to essentially duplicating the sampling effort at seven sites, in all measured parameters. This is prohibitively costly to do and hence an alternative method of documenting variance in personnel bias is presented.

At 10% of the streams sites (three per park total), efforts of the crew members should be duplicated (e.g., the Field Crew Leader does a discharge measurement, followed by another crew member). Another example is in repeating alkalinity measurements by both crew members. Over the life of the protocol, this should result in a gradually building body of knowledge about the degree of variation caused by different personnel. It is recommended that the annual reports (SOP #22: Data Reporting and Analysis) include the results of this duplication of efforts as an evaluation of protocol success. If large variation between personnel is detected, this should trigger an evaluation of the parameter that the variation is large in. What thresholds of variation

and technique of measuring variation might precipitate a re-evaluation will be dependent upon the parameter. Using alkalinity as an example, repeat measurements by different personnel that yield values of 20 mg/L and 23 mg/L are relatively very close and would generally be recognized as being in the neighborhood of low alkalinity, but the percent relative difference is actually somewhat high: (20 - 23)/20 = 15%. Percent relative difference is a commonly used measure for assessing QA/QC goals, but in this case, other techniques such as coefficients of variation may be more applicable. Hence, it is recommended that each parameter be evaluated by the Project Lead.

When repeating measurements at 10% of the stream sites, the repetition does not have to occur at the same sites. So alkalinity could be repeated at stream sites numbered 1, 3, 5, and 7, whereas the habitat could be repeated at stream sites numbered 2, 5, 9, and 12. The determination of what parameters to be repeated at which stream site should be a random determination by the Project Lead prior to the field season, although on the ground changes may be made so that sites can still sampled in a single day. Information for the field crews should be included in the site information folder (for example, see Appendix G). This will spread the repeat effort out across the sampling frame, so that the increased time for any one site is not beyond the logistical ability of the field crew. Care should be taken to document on the field forms and in the database which measurements are the duplicate.

It should also be stressed that control of personnel bias is done through strict, exact SOP adherence and through training. Furthermore, it is valuable to limit interpretation to measures that are less likely to be affected by personnel differences. For example, macroinvertebrate abundances can be highly affected by personnel experience; however, relative abundance or Presence/Absence measures are less affected by personnel. Hence, inferences to site or temporal impacts should be evaluated at the more robust measures first.

Change in Equipment

When possible, replacement equipment should match the original equipment. Specifications of replacement gear should also match the original equipment. If being replaced because of planned obsolescence, the new equipment should be tested against the old equipment to establish comparability.

Change in Contract Laboratories

Bias from different laboratories is a source of error that is under the control of the Project Lead. Generally, a change in labs used should have solid rationales, which include (but are not limited to), the following reasons:

(1) Lab shuts down or otherwise closes.
(2) Dramatic price increase so that the Network cannot meet sample completeness goals.
(3) Lab consistently fails to meet agreed upon deadlines for data delivery.
(4) Laboratory internal QA/QC procedures change or their methodology changes.

It is the responsibility of the Project Lead to stay in communication with contracting officers and laboratory managers to ensure that laboratories meet the standards and that lab continuity is maintained where and when appropriate.

SOP #19: Quality Assurance Project Plan (continued).

Programmatic Elements of the QAPP

The following covers aspects of the day to day data acquisition and data generation that must be followed to meet the requirements of this QAPP.

1. Sampling Methods

All data generated or acquired through this protocol must adhere to all SOPs. Crews should be trained and training documented to demonstrate that this aspect has been met.

2. Sample Handling and Custody

Certain basic requirements concerning filter choice, holding container, storage method, and storage time (Table 1). Biotic samples are invertebrates require no special handling or holding time, except to be preserved in 95% Ethanol. Chains of custody (Appendix F) must be maintained and logged to record the transfer and shipment of samples.

3. Measurement Quality Objectives and Reporting Limits

Measurement quality objectives (MQOs) are a set of attributes (e.g., precision, bias, sensitivity, detection limits, etc.) that determine whether or not a test result is accepted (Table 2). Reporting limits determine at what level an analyte test can be reported at a quantifiable level, or whether or not the proper determination is "detected, below quantifiable level" or "non-detect." Note that many labs are able to provide a "quantified" number for analytes below these levels. For analytical and exploratory purpose, the Project Lead may choose to use these numbers for calculations. However, for reporting and regulatory purposes, the reporting limits *must be adhered to*. Here, any value under "Method Detection Limit" must be reported as "Non-Detect" and values between "Minimum Level of Quantification" and "Method Detection Limit" must be reported as "Detected, below quantification limits," when the data are transferred to NPStoret and the EPA Storet. Internal databases will retain the raw value as reported by the lab, even if the values are below the Minimum Level of Quantification.

Table 1. Required sampling specifics to meet necessary goals of data quality for water chemistry analytes.

Analytes	Units	Required container (see SOP#1)	Required filter	Required sample volume	Required preservation	Maximum holding time
Filtered water sample:						
Anions (Cl, SO_4)	mg/L	Acid washed, HDPE	Whatman GF/C (1822-047)	50 ml	Frozen (-18°C)	28 days
Cations (Na, Ca, K, Mg)	mg/L	Acid washed, HDPE	Whatman GF/C (1822-047)	50 ml	Frozen (-18°C)	28 days
Total nitrogen	mg/L	Acid washed, HDPE	NA	100 ml	Frozen (-18°C)	28 days
Total phosphorous	mg/L	Acid washed, HDPE	NA	100ml	Frozen (-18°C)	28 days
			Total:	300 ml		

SOP #19: Quality Assurance Project Plan (continued).

Table 2. Required sampling specifics to meet necessary goals of data quality for water chemistry analytes (continued).

Analytes	Units	Required container (see SOP #1)	Required filter	Required sample volume	Required preservation	Maximum holding time
Other:						
Chlorophyll	mg/L	HDPE (storage for filter)	Millipore mixed Cellulose ester membrane (HAWP 047-00)	500 ml filtered	Frozen (-18°C)	28 days
Dissolved organic carbon	mg/L	Acid washed, furnace fired Amber Glass container	Whatman GF/F (1825-047)	60 ml	Refrigerated at 4°C	28 days

Table 3. Minimal Measurement Quality Objectives and reporting limits for chemical parameters necessary to be met by contract laboratories.

Parameter	Method Detection Limit (mg/L)	Minimum Level of Quantification (mg/L)	Precision (± mg/L)
Calcium	0.06	0.19	0.06
Chloride	0.01	0.03	0.01
Dissolved Organic Carbon	0.05	0.16	0.05
Total nitrogen	0.01	0.032	0.01
Total phosphorous	0.002	0.003	0.002
Magnesium	0.02	0.06	0.02
Potassium	0.03	0.1	0.03
Sodium	0.01	0.03	0.01
Sulfate	0.02	0.06	0.02

4. Chemical Laboratory Quality Controls

Chemical contract laboratories are required to meet certain responsibilities for assuring quality control. For this protocol, the required controls are:

a) Instrument calibration prior to initiating analysis run (three to six NIST traceable standards).
b) Standards analyzed every 10 samples.
c) Detection limit standard run at least once in analysis run.
d) A minimum of 10% of the samples must be duplicated (lab duplicate, not field duplicate).
e) Field duplicates should be run (responsibility of Project Lead to provide).
f) Periodic blanks should be run.

Filter blanks and bottle blanks should be included in the analysis. It is the responsibility of the Project Lead to have the field crew collect a minimum of four field duplicates during the field season. These are samples handled identical in every way to the original sample, but they are used to generate measures of precision for the analyses. The Project Lead must also provide bottle blanks (Acid-washed bottles filled with deionized water). Likewise, deionized water should be filtered using the protocol and then treated as samples. The purpose of these bottle and filter blanks is to ensure that there are no contamination sources in the bottle or filter preporation that may jeopardize results. Blanks, either filter or bottle, will be tracked in the database.

When duplicates or standards are greater than 10% deviant, the instruments should be recalibrated and the analyses repeated back to the last successful standard check.

Laboratories not conforming to these criteria should be rejected by the program during the initial contracting period.

5. Taxonomic Laboratory Quality Controls

Taxonomic laboratories should employ and use only taxonomists certified by the North American Benthological Society (NABS, www.benthos.org; an international scientific organization of aquatic ecologists) as trained taxonomists. There is currently no certification program for zooplankton taxonomists. For zooplankton taxonomists, a taxonomist C.V. or resume should be obtained and kept on record by the Project Lead as proof of taxonomic ability. This will standardize the knowledge so that invertebrate identifications can be assured as being correct. Taxonomic laboratories must also either: (1) complete a voucher collection of each taxon identified, or (2) retain and curate the portion of sample sorted and identified. Either of these will allow for double checking taxonomic accuracy by later project leaders or subject area experts. As the budget allows, periodic consideration should be made to sending taxonomic specimens to secondary labs to confirm taxonomic identifications.

6. Data Entry and Data Management

Data entry and data management are covered in SOP #4: Data Entry, but QA/QC procedures are described here.

Although the Klamath Network has opted to develop and maintain a Network-specific database, the database meets the requirement of the WRD to be consistent with NPS EDD (electronic data deliverables), so that all data will interface with both NPStoret and EPA STORET. This standardization and standard metadata requirements improve comparability of long-term datasets.

Prior to leaving the site, hardcopy datasheets, electronic data entry forms, specimen labels, and data logger data will be reviewed by the field crew before leaving the site. If possible, both field crew members should review all of the data. Field crew members should utilize built-in database utilities and their personal knowledge of the data to ensure all fields on the forms are complete and the entered data are logical and appropriate.

Data collected using datasheets during the project will be entered no less frequently than once a week. This will ensure (a) that details about the sampling event are still recent for the data recorders and (b) that the Project Lead can check on the progress of the data entry and associated tasks earlier rather than later. Hence, mistakes either in data collection or entry can be caught in time to rectify the problem.

Checks done by the Project Lead, both during the field season and at the end of the season include the following:

- Completeness checks - Are all forms in the database entered?
- Double checks - Are the entered values the same as on the datasheet? (A minimum of 20% should be checked; if there are a significant number of errors, 100% of the values should be checked.)

SOP #19: Quality Assurance Project Plan (continued).

- Outlier checks - Are there any values that are outside the normal range of variability (such as measurement or natural variability, etc), that might be suspect? Full details on how to deal with outliers is presented in Irwin (2008).
- Data flagging - Are there any values that have been flagged by the crew? The Project Lead must determine if to accept, reject, or redo a flagged value, including having the field crew return to the habitat for re-measurement.
- Formal data verification - Verification of the data involves evaluating the correctness, conformance, compliance, and completeness of the entire dataset against the methods or procedures of the protocol (SWAMP 2008). Verification should be done on both field data (including field chemical analyses) and laboratory data (chemical, invertebrate, and zooplankton).
    - Data verification includes:
        - Visual review at data entry – The technician verifies each value during input. Errors are corrected immediately.
        - Visual review after data entry – After entry, data are printed out and compared to original hardcopy sheets.
        - Duplicate data entry – Randomly selected site data are entered as normal but are duplicate records. Although time consuming in that it repeats data entry efforts, this gives an estimate of the data entry accuracy.
        - Review – It is the Project Lead's responsibility to review a subset of records to insure that they are identical to the hardcopy datasheets.
        - For the duplicate data and review, the minimum number is 20% of the sites (approximately eight total sites).
- Formal data validation - After verification, the Project Lead reviews them against all criteria in the protocol, especially the QAPP criteria (e.g., holding times, laboratory duplicates, completeness goals, reporting limits). After successful validation, the Data Manager can send the data onto WRD for incorporation into NPStoret.
    - Data Validation includes:
        - Data entry programming steps – The Project Lead, along with the Network Data Manager, will program steps designed to prevent errors. For example, maximum reach length will not allow an entry over 500 meters. This is an example of a mistake that might occur if the technician accidentally enters "2000" instead of "200."
        - Outlier detection and review (see Irwin 2008) – Statistical review and graphical display will be used to detect outliers (unusually extreme values of a variable outside the range of normal values). In outlier review, it is important to realize that not all extreme values represent errors, but can reflect the real variation of the data in nature. Generally, outliers that cannot be ascribed to error will be flagged and retained.
        - Review of "what makes sense" – The Field Crew Leader and Project Lead will compare and review the tabular data to confirm that everything "makes sense." Both should be intimately familiar with the types of data being collected and as such should be able to detect mistakes. GIS data will be plotted and confirmed to match the spatial locations.

SOP #19: Quality Assurance Project Plan (continued).

What Happens Now?

The QAPP is designed to identify, reduce, and, if possible, correct the data collected by the project. It is also designed to quantify the quality of the data, so that the precision of measurements of this protocol are to a known amount.

Data errors can never be entirely eliminated. Variation in the data due to collector, measurement, or equipment error can never be reduced to zero. If errors are so large so that data completeness goals are not met to the quality objectives of this protocol, corrective action should occur. The corrective steps should be commensurate with the severity of the errors. Possible steps include:

- Editing and documenting the error.
- "Re-dos," when possible (e.g., re-entering all the data).
- Revising specific SOPs.
- Increasing the training period.
- Eliminating SOPs with large, uncontrollable variation.
- Changing contract laboratories.
- Programmatic review.

Literature Cited

Irwin, R. J. 2008. Draft part B lite QA/QC review checklist for aquatic vital sign monitoring protocols and SOPs. National Park Service, Water Resources Division. Fort Collins, Colorado. Online. (http://www.nature.nps.gov/water/Vital_Signs_Guidance/Guidance_Documents/PartBLite.pdf.) Accessed 3 December 2009.

Surface Water Ambient Monitoring Program. 2008. Quality assurance project plan. California State Water Resources Control Board. Available online. (http://www.swrcb.ca.gov/water_issues/programs/swamp/docs/qapp/swamp_qapp_master090108a.pdf.) Accessed 3 December 2009.